

# Identification, Sequence, and Transcriptional Mapping of the Major Capsid Protein Gene of the Baculovirus *Autographa californica* Nuclear Polyhedrosis Virus

SUZANNE M. THIEM AND LOIS K. MILLER\*

*Departments of Genetics and Entomology, University of Georgia, Athens, Georgia 30602*

Received 14 November 1988/Accepted 12 January 1989

The gene encoding the major capsid protein of the baculovirus *Autographa californica* nuclear polyhedrosis virus (AcMNPV) was identified, sequenced, and transcriptionally mapped. The location of the gene was determined by immunological screening of an expression library of AcMNPV open reading frame- $\beta$ -galactosidase fusions with an antibody raised to virus structural proteins. The DNA sequence of the corresponding region, which mapped within 56.6 and 58.0 map units on the AcMNPV genome, revealed a 1,040-base-pair open reading frame capable of encoding a 39-kilodalton polypeptide. The identity of the polypeptide was determined by Western blot (immunoblot) analysis of purified empty capsids with an antibody raised to the capsid- $\beta$ -galactosidase fusion protein. The identity of the peptide encoded by the gene was confirmed by immunoprecipitation of an in vitro translation product with RNA selected by hybridization to DNA sequences from the coding region of the gene. Transcripts of the capsid gene were analyzed by Northern (RNA) blots and mapped by nuclease protection and primer extension analysis. The capsid gene is transcribed maximally at 12 and 24 h postinfection but not in the presence of cycloheximide, a protein synthesis inhibitor, or aphidicolin, a viral DNA synthesis inhibitor, and is therefore classified as a late gene. The gene is transcribed in a counterclockwise direction with respect to the circular map. There are three transcriptional start sites, all containing the  $\hat{G}$ TAAG consensus sequence found at the start site of all late AcMNPV genes.

*Autographa californica* nuclear polyhedrosis virus (AcMNPV) belongs to the family *Baculoviridae* and serves as a model system for the study of the molecular biology of baculoviruses (reviewed in references 8 and 13). Baculoviruses are used as biological pesticides (reviewed in reference 14) and have gained wide use recently as a helper-independent expression vector for foreign genes (22, 25, 26). AcMNPV has a rod-shaped capsid containing a circular double-stranded DNA genome of approximately 128 kilobase. Baculovirus nucleocapsids consist of the DNA condensed with a basic protamine-like core protein within the capsid (40). AcMNPV replicates in the nuclei of permissive insect cells and produces two morphological forms: budded virus (BV) and occluded virus (OV). BV consist of single enveloped nucleocapsids which bud from cells beginning by 10 to 12 h postinfection (p.i.) and serve to spread infection among cells within an individual insect. OV are formed in the nucleus and are composed of enveloped nucleocapsids which are enclosed within a crystalline protein matrix beginning 18 to 24 h p.i. The OV serve to spread infection among insects by an oral route.

The infection process is temporally regulated, with at least three phases of gene expression—early, late, and very late (reviewed in references 7 and 10). Late genes include those involved in virus assembly, while very late genes are involved specifically in the occlusion process. Transcriptional analysis indicates that the various temporal classes of genes are dispersed throughout the genome. Multiple overlapping transcripts with coterminal 5' or 3' ends are a common transcriptional motif. There is no evidence of splicing of late transcripts.

Few AcMNPV genes encoding specific products have been identified. Two occlusion-specific genes, polyhedrin (1,

17) and p10 (19), have been located, mapped, and sequenced. Recently, the gene encoding the arginine-rich core protein found in AcMNPV nucleocapsids was located, sequenced, and transcriptionally mapped (47). The transcripts of polyhedrin, p10, and the 6.9-kilodalton (kDa) core protein gene all originate within the common sequence motif ATAAG. A 42-kDa peptide identified immunologically in both morphological forms of AcMNPV (34) has been identified as the major capsid protein (38). In this study, we determined the location, sequence, and transcriptional patterns of the AcMNPV capsid protein gene. A potential transcriptional regulatory mechanism is discussed.

(This research was done in part at the Department of Bacteriology and Biochemistry, University of Idaho, Moscow, in partial fulfillment of Ph.D. degree requirements [S.M.T.] )

## MATERIALS AND METHODS

**Virus and cells.** The virus used for all studies was AcMNPV L-1 (21). OV was isolated from *Trichoplusia ni* larvae infected per os at the fourth instar. AcMNPV was propagated in *Spodoptera frugiperda* IPLB-SF-21 (SF-21) cells (42) maintained in TC100 medium (GIBCO Laboratories, Grand Island, N.Y.) supplemented with 0.26% tryptose broth, 0.6  $\mu$ g of amphotericin B per ml, 0.06 mg of penicillin G per ml, 0.27 mg of streptomycin sulfate per ml, and 10% fetal bovine serum. BV was prepared by infecting subconfluent monolayers of cells ( $10^7$  cells per 100-mm plate) with a multiplicity of 0.5 PFU per cell. For time courses, monolayers of cells ( $10^7$  cells per 100-mm plate or  $10^6$  cells per 35-mm plate) were inoculated with passage 2 AcMNPV at a multiplicity of 20 PFU per cell. The inoculum was removed after adsorption for 1 h at room temperature. Time zero was defined as the time when the inoculum was removed and incubation at 27°C was initiated. To block protein synthesis,

\* Corresponding author.

we used cycloheximide at 100  $\mu\text{g/ml}$  throughout the time course beginning with a 30-min pretreatment prior to the addition of inoculum. To block DNA replication, we added aphidicolin to medium at 5  $\mu\text{g/ml}$  following the adsorption period (33).

**Construction and screening of an ORF library.** A library of potential AcMNPV protein-coding regions was constructed by cloning random AcMNPV DNA fragments into a plasmid, pJS413 (45) (Molecular Genetics, Inc., Minnetonka, Minn.), which facilitates the selection of DNA fragments containing open reading frames (ORFs). Plasmid pJS413 contains the ribosome-binding site and  $\text{NH}_2$  terminus of the *cro* gene from bacteriophage lambda and the *lacZ* gene from *Escherichia coli* under the control of the *lac* operator and promoter. *cro* and *lacZ* are separated by a multiple cloning site which places the two genes out of translational reading frame with respect to each other, resulting in a  $\text{Lac}^-$  phenotype. A DNA fragment containing an ORF results in the production of a trihybrid fusion protein (*cro*-ORF-*lacZ*) and a  $\text{Lac}^+$  phenotype if it is inserted into the multiple cloning site such that the translational reading frames of *cro* and *lacZ* are placed back in phase. Random blunt-ended fragments of viral DNA were prepared by sonication and T4 DNA polymerase repair essentially by the method of Deininger (6). Fragments 300 to 600 base pairs (bp) long were selected and inserted into the *Sma*I site of pJS413, which was used to transform *E. coli* NF1829, a strain which overproduces *lac* repressor (43). Clones expressing  $\beta$ -galactosidase were selected and screened for the presence of viral DNA by colony hybridization by standard methods (23). The resulting library of clones expressing AcMNPV ORF- $\beta$ -galactosidase fusions was screened immunologically with antiserum against virus structural proteins by the method of Helfman et al. (15).

**Antibody production.** Antibodies were raised in rabbits by administration of 300 to 500  $\mu\text{g}$  of antigen emulsified in Freund complete adjuvant for initial injections and Freund incomplete adjuvant for subsequent injections by intradermal (41) and subcutaneous routes. Antiserum to virus structural proteins was raised against BV isolated from tissue culture. Antiserum to the selected AcMNPV ORF- $\beta$ -galactosidase fusion protein CG-4- $\beta$ -galactosidase was raised against an immunoaffinity column-purified fusion protein. Fusion protein production was induced in log-phase *E. coli* NF1829 cultures harboring the plasmid of interest by the addition of isopropyl- $\beta$ -D-thiogalactopyranoside (Sigma Chemical Co., St. Louis, Mo.) to 1 mM. Bacterial lysates were prepared, and the fusion proteins were purified on a Protosorb-*lacZ* immunoaffinity column as described in Promega Biotec Technical Bulletin 026 (Promega Biotec, Madison, Wis.).

**Isolation and purification of virus components.** AcMNPV extracellular enveloped nucleocapsids, BV, were purified from tissue culture supernatants by centrifugation for 30 min at  $40,000 \times g$  through a 10 to 50% (wt/wt) sucrose gradient in 0.01 M Tris hydrochloride (pH 7.5). The band of enveloped nucleocapsids was removed, diluted, and pelleted by centrifugation for 1 h at  $80,000 \times g$ . Enveloped nucleocapsids from OV (PDV), nucleocapsids, and capsids were purified from polyhedra by a modification of the method of Tween et al. (40). Enveloped nucleocapsids were released from polyhedra by treatment with 0.1 M sodium carbonate for 1 h at room temperature. The solution was neutralized by the addition of 1 M Tris hydrochloride (pH 7.5). After removal of undisrupted polyhedra by centrifugation at  $7,500 \times g$  for 10 min, the released PDV was layered onto a 10 to 50%

(wt/wt) sucrose gradient in 0.01 M Tris hydrochloride (pH 7.5) and centrifuged for 30 min at  $40,000 \times g$ . Bands of PDV were removed, diluted in 0.01 M Tris hydrochloride (pH 7.5), and pelleted by centrifugation for 1 h at  $80,000 \times g$ . Detergent solutions used in the preparations of capsids and nucleocapsids included the protease inhibitors phenylmethylsulfonyl fluoride (1 mM), leupeptin (20  $\mu\text{M}$ ), and pepstatin (40  $\mu\text{M}$ ), all purchased from Fluka BioChemika (Ronkonkoma, N.Y.). To prepare nucleocapsids, we stirred PDV at room temperature for 1 h in 1% (vol/vol) Nonidet P-40–30 mM NaCl–0.01 M Tris hydrochloride (pH 7.5). Nucleocapsids were separated from solubilized envelope proteins by centrifugation through 10 to 50% (wt/wt) sucrose gradients in 0.01 M Tris hydrochloride (pH 7.5) for 30 min at  $40,000 \times g$ . The band of nucleocapsids was removed, diluted in 0.01 M Tris hydrochloride (pH 7.5), and pelleted by centrifugation for 1 h at  $85,000 \times g$ . Empty capsids were isolated by incubating nucleocapsids in 2% Nonidet P-40–0.01 M EDTA–1 M NaCl–0.01 M Tris hydrochloride (pH 8.5) for 12 h at  $37^\circ\text{C}$ . Empty capsids were banded on performed CsCl gradients (1.18 to 1.55  $\text{g/cm}^3$  in 0.01 M Tris hydrochloride [pH 8.5]) by centrifugation at  $150,000 \times g$  for 2 h. The visible empty capsid band at 1.33  $\text{g/cm}^3$  was removed, diluted in 0.01 M Tris hydrochloride (pH 7.5), and pelleted by centrifugation at  $85,000 \times g$  for 30 min.

**Polyacrylamide gel electrophoresis and Western immunoblotting.** Viral proteins were analyzed by electrophoresis on sodium dodecyl sulfate (SDS)-10% polyacrylamide slab gels by the system of Laemmli (20). Samples were run in duplicate on the same gel. Following electrophoresis, the gel was divided in half. The proteins in one half were stained with Coomassie brilliant blue G, and the proteins in the other half of the gel were electrophoretically transferred to a nitrocellulose filter (39). The filter was probed immunologically with the antiserum raised against the AcMNPV structural ORF- $\beta$ -galactosidase fusion protein CG-4- $\beta$ -galactosidase. The immunopositive proteins were visualized with alkaline phosphatase-conjugated goat anti-rabbit immunoglobulin G and Nitro Blue Tetrazolium (Sigma) by the method of Blake et al. (3) except that 0.15 M Tris hydrochloride (pH 9.6) was substituted for the barbital buffer.

**Sequencing strategy.** The location of the ORF was determined by probing Southern blots of restricted AcMNPV DNA with radiolabeled plasmid DNA carrying the AcMNPV ORF- $\beta$ -galactosidase fusion (pJCG-4). Plasmids pSTCHX3 and pSTCHX3M were constructed by cloning the AcMNPV *Hind*III-C-*Xho*I-A fragment spanning 53.5 to 58.9 map units (m.u.) into the *Hind*III and *Xho*I sites of Bluescript KS(+) and KS(−) cloning vectors (Stratagene, San Diego, Calif.). To sequence the ORF, the AcMNPV fragment was excised from pJCG-4 by cutting, with *Bam*HI and *Bgl*II, sites flanking the *Sma*I ORF insertion site and inserted into the *Bam*HI site of Bluescript KS(+). A series of overlapping deletions were made in the cloned AcMNPV fragments with exonuclease III (16) and mung bean nuclease. Single-stranded deletion clones were sequenced by the dideoxyribonucleotide-chain termination method (35), using a modified T7 DNA polymerase and a Sequenase kit purchased from U.S. Biochemical Corp. (Cleveland, Ohio). The deduced capsid protein sequence was compared with the National Biomedical Research Foundation data base (release 10.0) and translated sequences from Genbank (release 55), using the IBI-Pustell DNA and protein analysis programs.

**Transcriptional mapping.** Monolayers of AcMNPV-infected SF-21 cells were harvested at various times p.i., and total cell RNA was isolated by the guanidinium isothiocya-

nate method (5). Northern (RNA) blot analysis was performed with 20  $\mu$ g of total RNA per lane. RNA was denatured by glyoxalation (28), fractionated on 1% agarose gels, and transferred to Zeta-Probe nylon membranes (Bio-Rad, Laboratories, Richmond, Calif.). The blots were probed with strand-specific RNA probes transcribed by T3 and T7 RNA polymerases from the *ScaI-NarI* fragment (57.1 to 57.6 m.u.) of the capsid ORF cloned into Bluescript KS(+) (pSTNSP).

The 5' and 3' ends of the transcripts were mapped by nuclease protection analysis (44), with either mung bean or S1 nuclease. Probes were generated by radiolabeling the appropriate recombinant plasmid at a single restriction site with either T4 polynucleotide kinase for 5' ends or T4 DNA polymerase for 3' ends. The plasmids were cleaved with a second enzyme and purified on agarose gels to yield probes labeled exclusively at one end. Hybridizations were conducted at 45°C in sodium trichloroacetic acid buffer (27) or at 49°C in 80% formamide–40 mM PIPES [piperazine-*N,N'*-bis(2-ethanesulfonic acid)] (pH 6.4)–0.4 M NaCl–1 mM EDTA. Nuclease-resistant fragments were analyzed on 6% polyacrylamide–7 M urea sequencing gels.

For primer extension analysis, a 15-bp oligonucleotide primer corresponding to the *NarI* site used in nuclease protection studies was synthesized and labeled at its 5' end with T4 polynucleotide kinase. The primer was annealed to total infected cell RNA isolated at 12 h p.i. under the same conditions used for nuclease protection analysis. The primer was extended with cloned Moloney murine leukemia virus reverse transcriptase (Bethesda Research Laboratories, Inc., Gaithersburg, Md.). A sequencing ladder was generated by using the labeled oligonucleotide to prime a dideoxynucleotide-chain termination sequencing reaction from a single-stranded DNA spanning the region.

**Hybrid selection and in vitro translation.** RNA homologous to the capsid ORF was isolated by the method of Esche and Siegmund (9) as modified by Friesen and Miller (11). A single-stranded phage DNA (30  $\mu$ g) generated from a Bluescript KS(–) plasmid carrying the capsid ORF *ScaI-NarI* fragment, pSTNSM, was immobilized on a nitrocellulose filter. Hybridization was conducted at 42°C in 50% formamide–10 mM PIPES (pH 6.4)–0.4 M NaCl–1 mM EDTA for 12 h with 1.5 mg of total RNA isolated from SF-21 cells 12 h p.i. After hybridization, the filters were washed extensively in 1× SSC (0.15 M NaCl and 0.015 M trisodium citrate, pH 7.0)–0.1% SDS–2 mM EDTA, and the selected RNA was eluted by boiling for 1 min in 1 mM EDTA containing 10  $\mu$ g of yeast tRNA (Sigma). In vitro translation was performed at 30°C for 1 h in a rabbit reticulocyte system (Promega Biotec) in the presence of 50  $\mu$ Ci of [ $^{35}$ S]methionine (1,200 Ci/mM; Dupont, NEN Research Products, Boston, Mass.) by the method of Pelham and Jackson (30). Proteins from infected cells were radiolabeled by incubating previously infected cell monolayers ( $10^6$  cells per 35-mm plate) in methionine-deficient growth medium containing 200  $\mu$ Ci of [ $^{35}$ S]methionine per ml. Following 1-h incubations at 6, 12, and 24 h p.i., cells were collected, washed with phosphate-buffered saline (31), and lysed in 1% Nonidet P-40–50 mM Tris hydrochloride (pH 8.0)–150 mM NaCl. Pulse-labeled proteins from infected cells and radiolabeled in vitro translation products were immunoprecipitated with anti-CG-4- $\beta$ -galactosidase antiserum and *Staphylococcus aureus* (Sigma) by the method of Kessler (18). Lysates were preprecipitated once with preimmune serum prior to incubation with immune serum. Infected cell proteins, in vitro translation products, and immu-

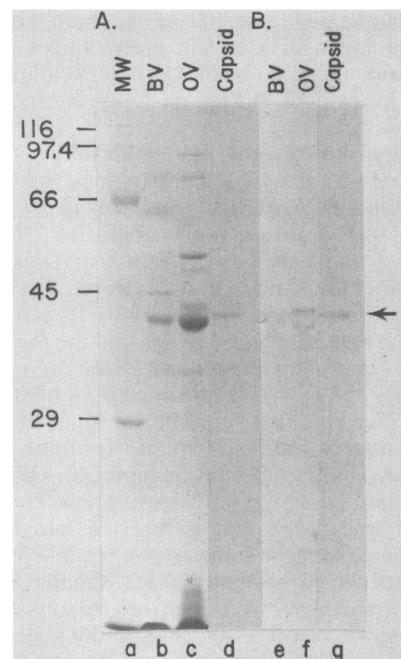


FIG. 1. SDS-polyacrylamide gel and Western blot analyses of AcMNPV structural proteins. Extracellular enveloped nucleocapsids (BV) (lanes b and e), enveloped nucleocapsids derived from OV (PDV) (lanes c and f), and capsids isolated from PDV-derived nucleocapsids (lanes d and g) were purified (see Materials and Methods). Samples were denatured, loaded in duplicate, and subjected to electrophoresis on an SDS-polyacrylamide gel. Half of the gel was stained with Coomassie brilliant blue G (A). The polypeptides in the other half were electrophoretically transferred to a nitrocellulose filter, incubated with anti-CG-4- $\beta$ -galactosidase antiserum, and immunologically stained (B). Molecular size standards (lane MW) in kilodaltons are shown on the left. The arrow indicates the capsid protein.

noprecipitates were analyzed by SDS-10% polyacrylamide gel electrophoresis followed by fluorography.

## RESULTS

**Identification of major capsid gene.** To identify specific AcMNPV structural genes, a library of plasmids carrying random AcMNPV ORF- $\beta$ -galactosidase gene fusions was constructed and immunologically screened with an antiserum raised against BV structural proteins. A colony producing an epitope recognized by the antiserum was identified by its strong reaction in the immunological screen of the library. The plasmid carried by this colony, pJCG-4, contained a 500-bp insert of AcMNPV mapping to the *HindIII-C-XhoI-A* region (53.5 to 58.9 m.u.) of AcMNPV (see below). The AcMNPV ORF- $\beta$ -galactosidase fusion protein produced from pJCG-4, CG-4- $\beta$ -galactosidase, was purified and used to raise antibodies to CG-4 epitopes.

To determine which virus structural protein was recognized by the anti-CG-4 antibody, the structural proteins of enveloped viruses from infected cell culture medium (BV) and pdv from infected insects (OV) were separated on SDS-polyacrylamide gels (Fig. 1, lanes b and c), blotted, and immunologically screened with anti-CG-4- $\beta$ -galactosidase antiserum (Fig. 1, lanes e and f). The antiserum reacted primarily with the major protein component of both BV and OV enveloped nucleocapsids (Fig. 1, lanes b and c), a

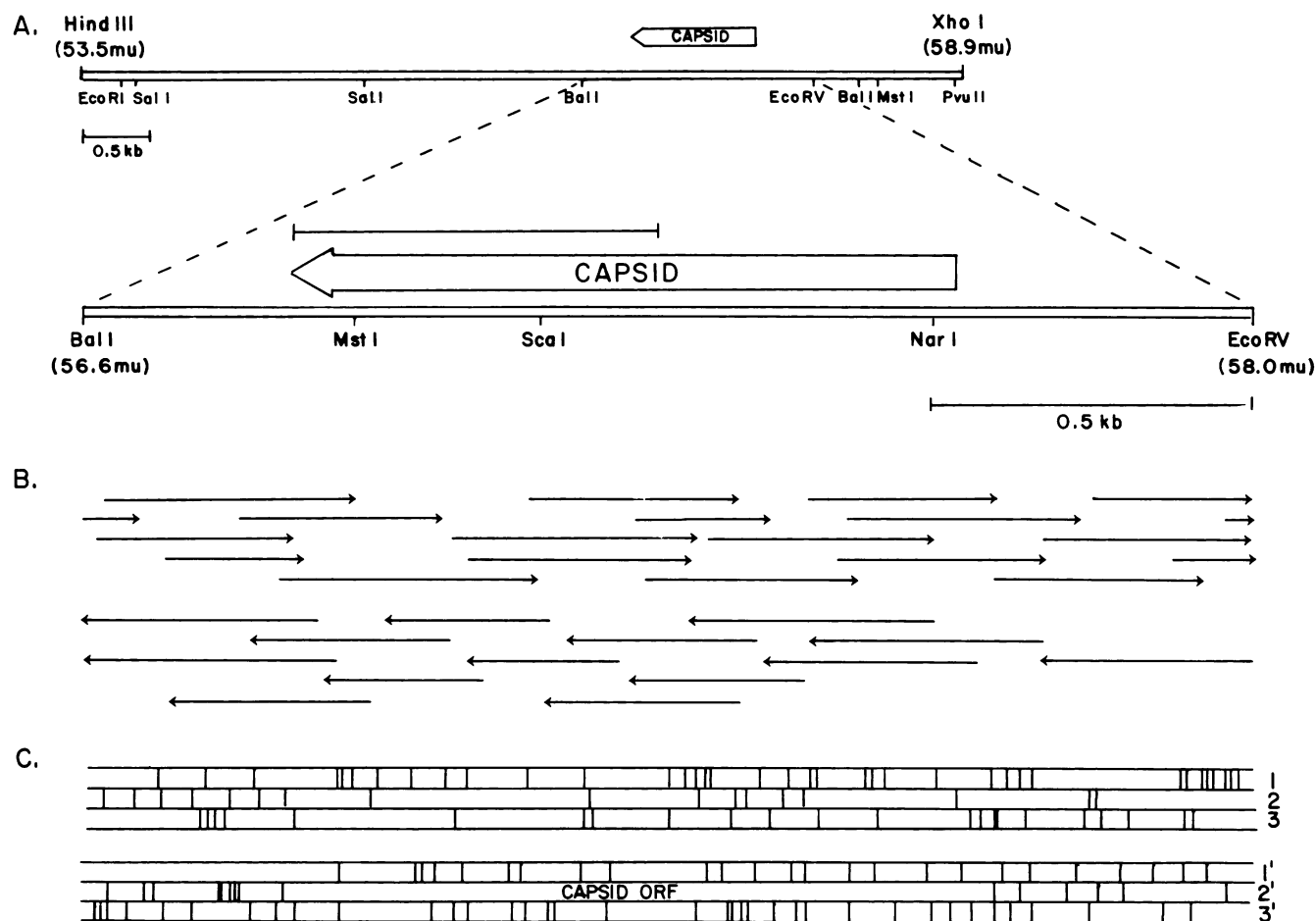


FIG. 2. Map, sequencing strategy, and ORF distribution of the AcMNPV *HindIII*-C-*XhoI*-A (53.5 to 58.9 m.u.) fragment. (A) Map of selected restriction endonuclease recognition sites of the *HindIII*-C-*XhoI*-A fragment. The *BamHI*-*EcoRV* region (56.6 to 58 m.u.) is expanded to show the location of the capsid ORF, depicted as an open arrow. The bar indicates the location of the 500-bp AcMNPV ORF carried by pJCG-4. This fragment was used to locate the capsid gene and produce the CG-4- $\beta$ -galactosidase fusion protein for immunological studies. kb, Kilobase. (B) Sequencing strategy for the region from 56.6 to 58 m.u. on the same scale as the expanded *BamHI*-*EcoRV* fragment. Individual deletion clones are depicted by arrows starting at the beginning of the insert and pointing in the direction the clone was sequenced. (C) Distribution of ORFs within the region from 56.6 to 58 m.u. Translation termination codons TAA, TAG, and TGA are depicted by vertical bars. The three reading frames for the top DNA strand are labeled 1, 2, and 3. The three reading frames for the bottom strand are labeled 1', 2', and 3'. The capsid ORF is located in reading frame 2'.

polypeptide of approximately 42 kDa, suggesting that the antiserum recognized the major capsid protein. Owing to the large quantities of the 42-kDa protein in enveloped nucleocapsid lanes (Fig. 1, lanes b, c, e, and f), the Western blot shows a bowing effect in the 42-kDa region. The smaller polypeptides reacting with the anti-CG-4- $\beta$ -galactosidase antiserum in the blot (lane f) appear to result from proteolytic cleavage of the capsid protein. During this work, we found it necessary to include substantial quantities of protease inhibitors in nucleocapsid preparations to minimize proteolytic cleavage of the capsid protein. Preimmune serum did not react with any viral proteins when tested on similar Western blots (data not shown).

The above results indicate that the anti-CG-4 antibody was recognizing the major capsid of both BV and OV. To confirm that the antiserum was recognizing the major capsid protein, empty capsids were purified and analyzed on SDS-polyacrylamide gels. The 42-kDa protein was the only visible component of purified preparations of empty capsids (Fig. 1, lane d), and this protein reacted with anti-CG-4- $\beta$ -galactosidase antiserum (Fig. 1, lane g). Thus, the protein recognized by

the anti-CG-4- $\beta$ -galactosidase antiserum is the AcMNPV major capsid protein.

**Mapping and sequencing the capsid gene.** The location of the capsid ORF was mapped by Southern blot analysis to the region of the AcMNPV physical map corresponding to restriction fragments *HindIII*-C and *XhoI*-A (53.5 to 58.9 m.u.). The AcMNPV *HindIII*-C-*XhoI*-A fragment (Fig. 2A) was cloned (pSTCHX3 and pSTCHX3M). Since there were few useful restriction sites within the region, the ORF was localized more precisely by hybridizing  $^{32}$ P-labeled pJCG-4 to Southern blots of a series of exonuclease III deletion clones of pSTCHX3 and pSTCHX3M. Once the location of the ORF was defined by the deletion clones, both strands of DNA from the region containing the capsid gene were sequenced with the appropriate deletion clones (Fig. 2B). The sequence revealed an ORF of 1,040 bp capable of encoding a 39-kDa polypeptide (Fig. 3) reading counter-clockwise with respect to the circular map of the virus. Several smaller ORFs reading in both directions were observed (Fig. 2C). When the deduced amino acid sequence of the capsid protein was screened against data bases, similar-

|                                                                                    |      |
|------------------------------------------------------------------------------------|------|
| GG ATA TCT TGT TCG CCA TCG TGG AAT CAA ATA GAT CAA TGT CAC TTT TCG AAA AAT ATA CAT | 62   |
| GTT CAA ATT TGA TTT CAA TTT TAT CGT GTT GGT AAA CGT ACA CTT TAA TTA TTT TAC TCA    | 122  |
| AGT TGT GCG AAA GAG TCT TGT AAG GCA GTT TGA TTT CTT TGC TTT CTC TCC ACA CCA ACG    | 182  |
| GCA CCA ACG CGT TGG TAT CTT TAG GCC AAT AAA CAA ATT TTT TGT GTT TGG AAT TAG TCT    | 242  |
| TTT TCA CGC TTG ATA TTA TGT TAT TGC AAG CGC TCT GAA TAG GTA TAC GAG TGC GAA AGC    | 302  |
| CGT TTT CGT CGT ACA AAT CGA AAT ATT GTG TGC CAG CGA ATA ATT AGG AAC AAT ATA AGA    | 362  |
| ATT TAA AAT TTT ATA CAA CAA ATC TTG GCT AAA ATT TAT TGA ATA AGA GAT TTC TTT CTC    | 422  |
| AAT CAC AAA ATC GCC GTA GTC CAT ATT TAT AAC GGC AAC AAT ATG GCG CTA GTG CCC GTG    | 482  |
| Met Ala Leu Val Pro Val                                                            |      |
| GGT ATG GCG CCG CGA CAA ATG AGA GTT AAT CGC TGC ATT TTC GCG TCC ATC GTG TCG TTC    | 542  |
| Gly Met Ala Pro Arg Gln Met Arg Val Asn Arg Cys Ile Phe Ala Ser Ile Val Ser Phe    |      |
| GAC GCG TGC ATA ACA TAC AAA TCG CCG TGT TCG CCC GAC GCG TAT CAT GAC GAT GGA TGG    | 602  |
| Asp Ala Cys Ile Thr Tyr Lys Ser Pro Cys Ser Pro Asp Ala Tyr His Asp Asp Gly Trp    |      |
| TTT ATT TGC AAC AAC CAC CTC ATA AAA CGT TTT AAA ATG TCA AAA ATG GTT TTG CCC ATT    | 662  |
| Phe Ile Cys Asn Asn His Leu Ile Lys Arg Phe Lys Met Ser Lys Met Val Leu Pro Ile    |      |
| TTC GAC GAA GAC GAC AAT CAA TTC AAA ATG ACG ATC GCT AGG CAT TTA GTT GGA AAT AAA    | 722  |
| Phe Asp Glu Asp Asp Asn Gln Phe Lys Met Thr Ile Ala Arg His Leu Val Gly Asn Lys    |      |
| GAA AGA GGT ATC AAG CGA ATT TTA ATT CCA AGC GCA ACC AAT TAC CAA GAC GTG TTT AAT    | 782  |
| Glu Arg Gly Ile Lys Arg Ile Leu Ile Pro Ser Ala Thr Asn Tyr Gln Asp Val Phe Asn    |      |
| CTA AAC AGT ATG ATG CAA GCC GAA CAG CTA ATC TTT CAT TTG ATA TAT AAC AAC GAA AAC    | 842  |
| Leu Asn Ser Met Met Gln Ala Glu Gln Leu Ile Phe His Leu Ile Tyr Asn Asn Glu Asn    |      |
| GCA GTT AAC ACT ATA TGC GAC AAT CTA AAA TAT ACC GAA GGT TTC ACA AGC AAC ACG CAA    | 902  |
| Ala Val Asn Thr Ile Cys Asp Asn Leu Lys Tyr Thr Glu Gly Phe Thr Ser Asn Thr Gln    |      |
| CGC GTT ATA CAC AGC GTT TAC GCA ACT ACA AAA AGC ATT CTG GAC ACC ACA AAC CCG AAC    | 962  |
| Arg Val Ile His Ser Val Tyr Ala Thr Thr Lys Ser Ile Leu Asp Thr Thr Asn Pro Asn    |      |
| ACG TTT TGT TCG CGG GTG TCG CGA GAC GAA TTG CGT TTC TTT GAC GTG ACC AAC GCC CGA    | 1022 |
| Thr Phe Cys Ser Arg Val Ser Arg Asp Glu Leu Arg Phe Phe Asp Val Thr Asn Ala Arg    |      |
| GCG CTT CGA GGC GGT GCT GGC GAT CAA TTA TTT AAC AAT TAC AGT GGA TTT TTG CAA AAT    | 1082 |
| Ala Leu Arg Gly Gly Ala Gly Asp Gln Leu Phe Asn Asn Tyr Ser Gly Phe Leu Gln Asn    |      |
| TTG ATT CGA CGC GCA GTA GCG CCC GAG TAC TTG CAA ATC GAC ACG GAG GAA TTG AGG TTT    | 1142 |
| Leu Ile Arg Arg Ala Val Ala Pro Glu Tyr Leu Gln Ile Asp Thr Glu Glu Leu Arg Phe    |      |
| AGA AAT TGC GCC ACG TGT ATA ATT GAC GAA ACG GGT CTG GTC GCG TCT GTG CCC GAC GGC    | 1202 |
| Arg Asn Cys Ala Thr Cys Ile Ile Asp Glu Thr Gly Leu Val Ala Ser Val Pro Asp Gly    |      |
| CCC GAG TTG TAC AAC CCG ATA AGA AGC AGT GAC ATT ATG AGA AGT CAA CCC AAT CGT TTG    | 1262 |
| Pro Glu Leu Tyr Asn Pro Ile Arg Ser Ser Asp Ile Met Arg Ser Gln Pro Asn Arg Leu    |      |
| CAA ATT AGA AAC GTT TTG AAA TTT GAA GGC GAC ACA CGT GAG CTG GAC AGA ACG CTT AGC    | 1322 |
| Gln Ile Arg Asn Val Leu Lys Phe Glu Gly Asp Thr Arg Glu Leu Asp Arg Thr Leu Ser    |      |
| GGA TAC GAA GAA TAC CCG ACG TAC GTT CCG CTG TTT TTG GGA TAC CAA ATA ATC AAT TCA    | 1382 |
| Gly Tyr Glu Glu Tyr Pro Thr Tyr Val Pro Leu Phe Leu Gly Tyr Gln Ile Ile Asn Ser    |      |
| GAA AAC AAC TTT TTG CGC AAC GAC TTT ATA CCA AGA GCA AAT CCT AAC GCT ACT CTG GGC    | 1442 |
| Glu Asn Asn Phe Leu Arg Asn Asp Phe Ile Pro Arg Ala Asn Pro Asn Ala Thr Leu Gly    |      |
| GGC GGC GCA GTG GCA GGT CCT GCG CCT GGT GTT GCA GGC GAA GCA GGT GGA GGA ATA GCC    | 1502 |
| Gly Gly Ala Val Ala Gly Pro Ala Pro Gly Val Ala Gly Glu Ala Gly Gly Gly Ile Ala    |      |
| GTC TAA AAA TGG AGT TTG TTA AAT TGC AAT GCA ACA TTT GTT TTT CGG TTG CAG AAA TTA    | 1562 |
| Val ---                                                                            |      |
| AAA ATT ATT TTC TGC AAC CAA TAG ACA GAT TGA CTA TAA TAC CCG TAT TAG AAC TAG ACA    | 1622 |
| CGT GCA AAC ATC AAT TAT GCT CAA TGT GTA TAC GCA AGA TTA GAA AAC GAA AAA AAG TAC    | 1682 |
| CAT GTC CCT TGT GTA GGG TTG AAA GTT TGC ATT TCA ACG TTT ACA GCG TAA ACC GAA ACG    | 1742 |
| TTG TGG ATG TGA TTA AAT GCA GCG CGT CGA GTG TCG CAC AAT GGA ATA AAA TAA ACG CTA    | 1802 |

ATT TTG ATG CGG CCT CTT TGG CCA

FIG. 3. Nucleotide sequence and predicted amino acid sequence of the AcMNPV major capsid gene. The last nucleotide in each line is numbered at the right. Transcriptional start sites ( $\hat{\Delta}$  TAAAG) are indicated by asterisks. The transcriptional initiation points at positions 141, 357, and 404 are designated by arrows. Potential polyadenylation signals (AATAAA) are underlined. An AATTAAA sequence (position 1558) at the 3' end of the short capsid protein gene transcript that may be utilized poorly as a polyadenylation signal is printed in bold face. Arrowheads indicate the nucleotides at positions 491 (*NarI*) and 1399 (*MstI*) that were radiolabeled to prepare probes used for transcriptional mapping.

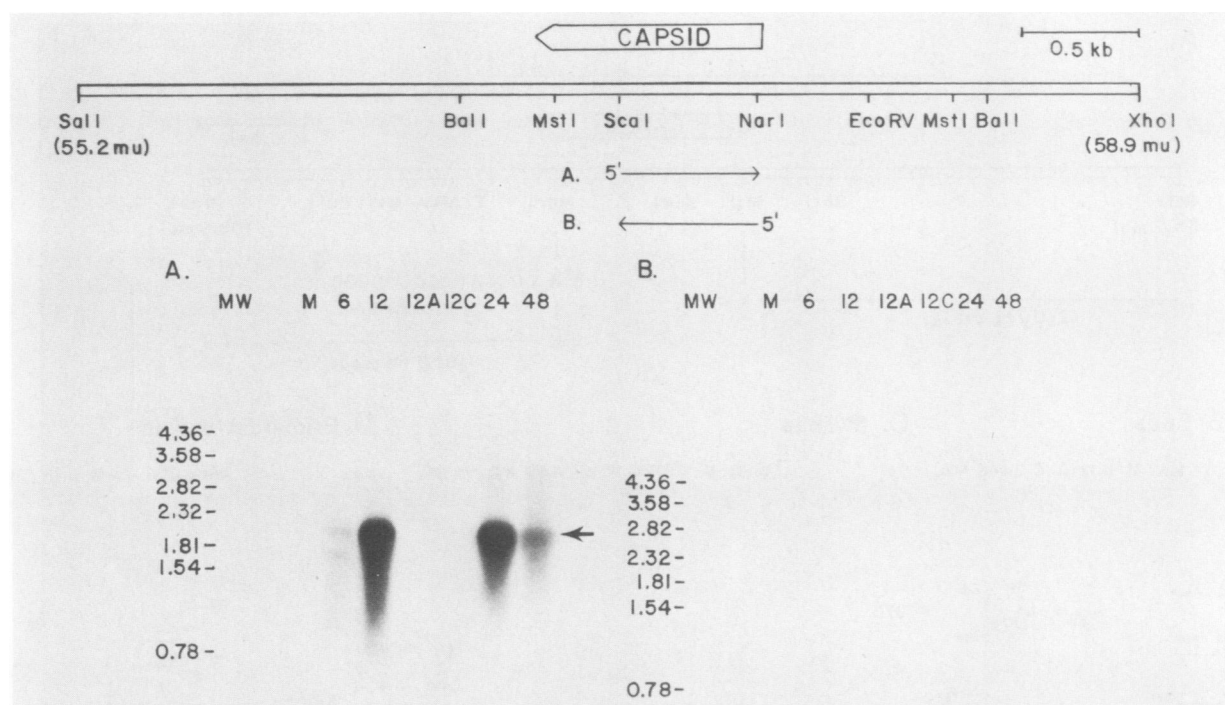


FIG. 4. Northern blot hybridization analysis of RNA homologous to strand-specific capsid probes. Total RNA from mock-infected cells (lane M), AcMNPV-infected cells collected 6, 12, 24, and 48 h p.i., and AcMNPV-infected cells incubated in the presence of aphidicolin (12A) or cycloheximide (12C) and collected 12 h p.i. was hybridized to a  $^{32}$ P-labeled 620-nucleotide antisense RNA (clockwise) probe (A) or a  $^{32}$ P-labeled 620-nucleotide sense RNA (counterclockwise) probe (B) of the capsid ORF transcribed from plasmid pSTNSP, containing the *ScaI*-*NarI* (57.1- to 57.6-m.u.) fragment of AcMNPV. The locations of the probes, depicted as arrows pointing 5' to 3', are shown in the diagram above the autoradiograms. Molecular size standards (MW) are indicated in kilobases (kb) on the left of each panel. The arrow in panel A indicates the position of the capsid transcript.

ities were found with short regions of three viral structural proteins. The greatest similarity noted was to a rotavirus capsid protein (4).

**Northern blot analysis.** Because AcMNPV genes can be transcribed from both DNA strands within the same region (11), we utilized strand-specific probes to monitor transcription from both strands by Northern blot analysis (Fig. 4). A probe complementary to the capsid ORF coding strand hybridized to a major 2.2-kilobase transcript expressed maximally at 12 and 24 h. This RNA was not expressed in infected cells incubated in the presence of aphidicolin or cycloheximide (Fig. 4A). The opposite strand (sense) probe did not hybridize to any transcripts (Fig. 4B), indicating that there is no detectable transcription in a clockwise direction through this region of the capsid ORF.

**Location of 5' and 3' ends of capsid transcript by nuclease protection and primer extension analysis.** The 3' ends of the RNA were mapped by an S1 nuclease protection assay with a probe 3' end labeled at the *MstI* site (Fig. 5A; nucleotide 1399 in Fig. 3). One major protected band of 975 nucleotides was observed at 12, 24, and 48 h p.i. (Fig. 5B). A minor band of 160 nucleotides was also observed at this time. Additional minor bands observed throughout the time course are probably artifacts resulting from incomplete nuclease digestion or nuclease cleavage within A+T-rich regions. A comparison of the DNA sequence with the minor 160-nucleotide protected fragment indicated either a polyadenylation or transcriptional termination site at or near an AATAAA sequence at nucleotide 1557 (Fig. 3), the only sequence resembling a consensus polyadenylation signal (2, 46) in this region. Although the 975-nucleotide protected fragment rep-

resented the major polyadenylation or termination site for late capsid transcripts, this protected fragment was also observed at 6 and 12 h p.i. in the presence of aphidicolin and cycloheximide, indicating that an early transcript initiates upstream from the *MstI* site. There are two AATAAA polyadenylation signals upstream from the termination site reflected by the 975-nucleotide fragment. The sequence of this region and the nature of this early transcript are the subject of another study (S. M. Thiem and L. K. Miller, submitted for publication).

To map the 5' ends of the capsid transcripts, we performed nuclease protection analysis on total infected cell RNA isolated at various times p.i. using a 1,650-nucleotide probe which was 5' end labeled at the *NarI* site (Fig. 5; nucleotide 491 in Fig. 3). RNA isolated at 12, 24, and 48 h p.i. protected primarily DNA fragments of 84, 132, and 348 nucleotides (Fig. 5C), although the intensity of the bands corresponding to the 132- and 348-nucleotide protected fragments was reduced at 48 h p.i. Protected fragments were not observed with RNA isolated at 6 h p.i. or at 12 h p.i. in the presence of aphidicolin or cycloheximide, indicating that capsid transcripts are expressed exclusively at late times in viral infection. The additional minor protected fragments are probably due to incomplete mung bean nuclease digestion of the single-stranded regions of the DNA.

For primer extension analysis, a 5'-end-labeled 15-nucleotide synthetic oligonucleotide corresponding to the *NarI* end of the nuclease protection probe (Fig. 5A) was annealed to RNA isolated at 12 h p.i. and extended with reverse transcriptase. Three extension products were synthesized (Fig. 5D) which corresponded exactly with the protected



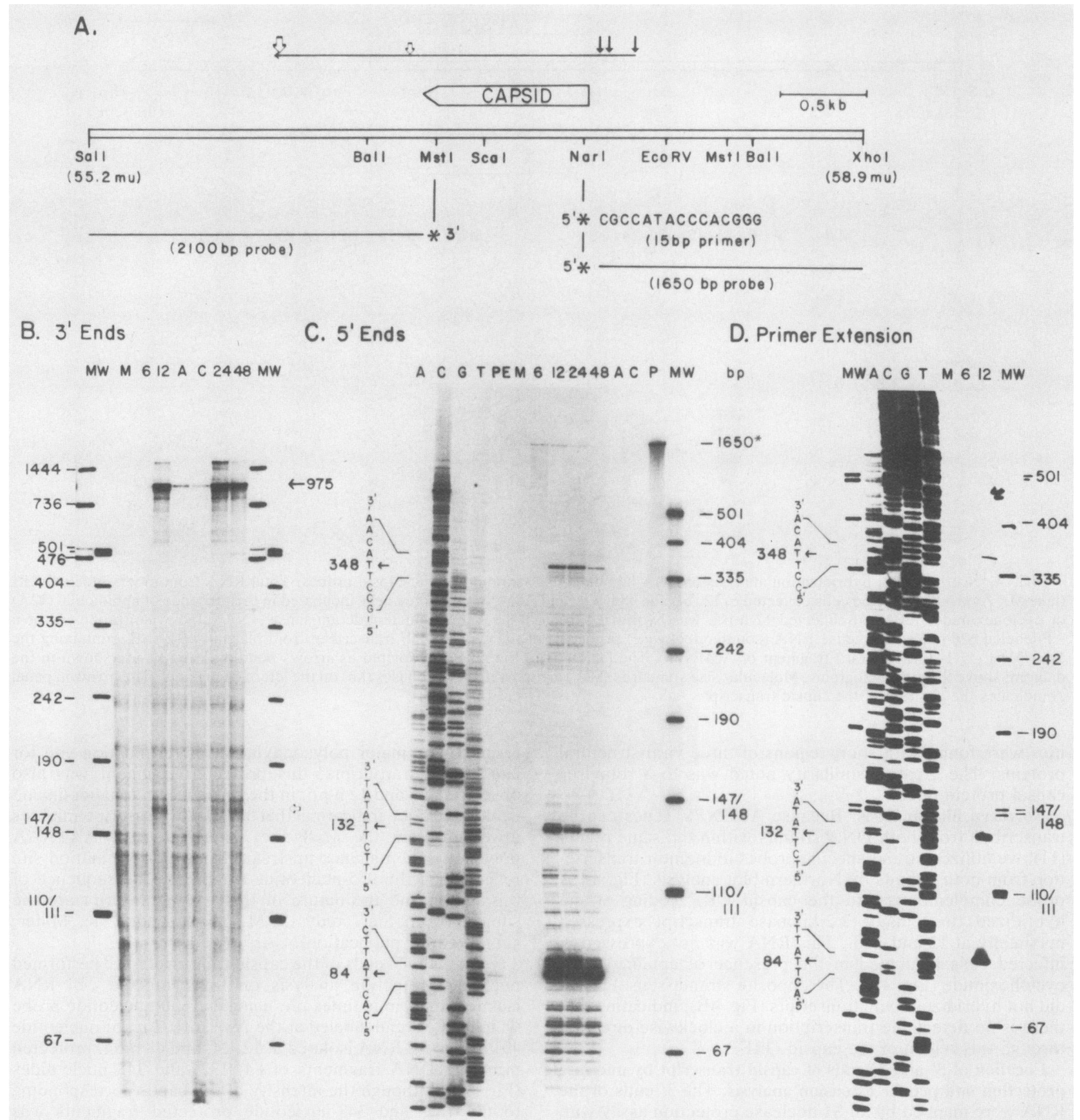


FIG. 5. Mapping the 3' and 5' ends of the capsid transcripts. (A) Locations of the probes and primer used to map the ends of the capsid transcripts relative to the capsid ORF. The size of each probe is shown below the line, and the radiolabeled ends are indicated by asterisks. At the top of the diagram, the location of the transcript is shown by an arrow, with the 3' ends indicated by open arrows and the 5' ends indicated by closed arrows. (B) S1 nuclease mapping of the 3' ends of the transcript. Total RNA from mock-infected cells (lane M), AcMNPV-infected cells at 6, 12, 24, and 48 h p.i., and AcMNPV-infected cells at 12 h p.i. in the presence of aphidicolin (lane A) or cycloheximide (lane C) was hybridized to the 2,100-bp 3' probe. The nuclease-resistant DNA-RNA hybrids were denatured and fractionated on a sequencing gel. The sizes (in nucleotides) of the molecular weight standards (lanes MW) and the probe are shown on the left. The predominant protected fragment of 975 bases is designated by the arrow on the right and by a large open arrow in panel A. A minor 160-base fragment is designated by an open arrow to the right of panel B and a small open arrow in panel A. (C) Mapping of the 5' ends of the transcript by mung bean nuclease digestion and primer extension. A sequencing ladder was made by dideoxynucleotide-chain termination sequencing of the region with the 5'-end-labeled oligonucleotide as a primer. For primer extension (lane PE), the 5'-end-labeled oligonucleotide was hybridized to 12-h RNA and extended with reverse transcriptase. For mung bean nuclease analysis, total RNA as described for panel B was hybridized to the 1,650-bp 5'-end-labeled probe and treated with mung bean nuclease. Primer extension products (lane PE) and nuclease-resistant DNA-RNA hybrids (lanes M, 6, 12, 24, 48, A, and C) were denatured and fractionated on a sequencing gel.

fragments observed in the nuclease protection assay (Fig. 5C). A sequencing ladder generated from an appropriate cloned DNA primed with the labeled oligomer revealed that all three transcripts started at the third nucleotide of an ATAAG or GTAAG sequence (Fig. 5C and D).

**In vitro translation of a hybrid selected mRNA corresponding to the major capsid protein.** To confirm that the ORF encoded a 42-kDa protein, we used a single-stranded DNA containing the *ScaI-NarI* fragment (57.1 to 57.6 m.u.), previously used as a probe for Northern blots (Fig. 4), to hybrid select mRNA isolated at 12 h p.i. The selected mRNA was translated in a cell-free protein synthesis system, and the products were analyzed by SDS-polyacrylamide gel electrophoresis. The translation product was a 42-kDa protein (Fig. 6, lane f). Immunoprecipitation of in vitro translation reactions showed that this protein is recognized by the anti-CG-4- $\beta$ -galactosidase antiserum and comigrates with the capsid protein immunoprecipitated from pulse-labeled infected cell lysates (Fig. 6, lanes d and e). This protein was absent in immunoprecipitates from mock-infected cells (Fig. 6, lane c).

Precipitated proteins from both in vivo and in vitro synthesis appeared as 42-kDa doublets in SDS-polyacrylamide gels (Fig. 6, lanes d and e). The fact that both in vivo and in vitro-derived proteins migrated as doublets suggests that the doublet is not a result of posttranslational modification unless proteolysis is involved. A more likely possibility is that protein synthesis could be initiated at an internal methionine codon like the herpes simplex virus thymidine kinase mRNA (24). There are three methionine codons within the first 40 bp of the major ORF.

## DISCUSSION

Through the use of an ORF selection-expression vector and immunological techniques, we identified the gene encoding the major capsid protein from AcMNPV. A molecular size of 39 kDa calculated from the predicted amino acid sequence agrees with the reported sizes of the capsid protein that range from 37 to 42 kDa (29, 34, 36–38). We demonstrated that we located the major capsid protein gene by (i) Western blot analysis of proteins from purified empty capsids, (ii) immunoprecipitation of pulse-labeled infected cell peptides, and (iii) analysis of products from in vitro translation hybrid-selected mRNA. We propose to call this gene *vp39* for virus structural protein 39.

We classify the capsid gene as a late gene based on the temporal appearance of transcripts and the effects of inhibitors on capsid gene transcription. Late viral transcripts are distinguished from early transcripts by a dependence on viral DNA replication. The capsid transcript was observed only in the absence of aphidicolin, a DNA synthesis inhibitor, and cycloheximide, a protein synthesis inhibitor. The reduction of capsid transcripts in the presence of cycloheximide indicates that virus-induced proteins are required for capsid gene transcription. This is consistent with a cascade model

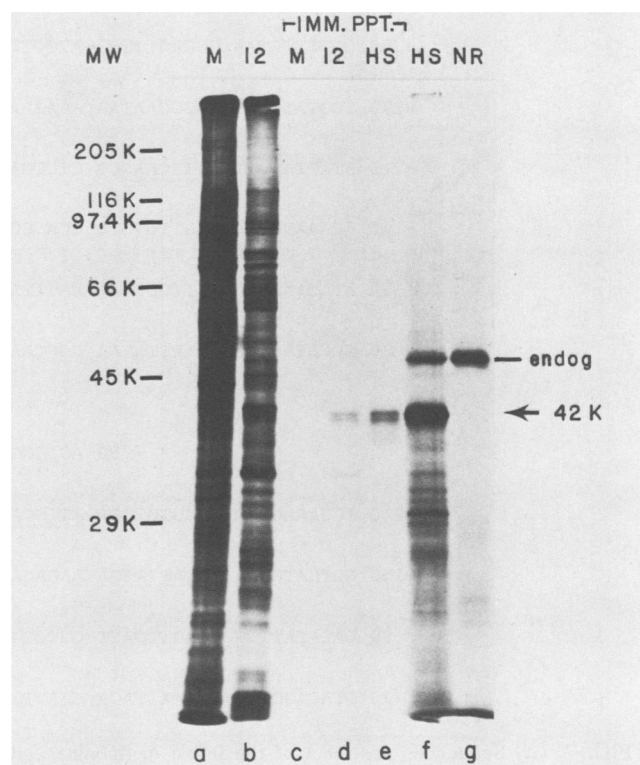


FIG. 6. Hybrid selection, in vitro translation, and immunoprecipitation of the capsid protein. Capsid RNA was selected from total virus RNA isolated 12 h p.i. with a strand-specific DNA clone of the capsid ORF immobilized on nitrocellulose and translated in an RNA-dependent rabbit reticulocyte translation system. The radio-labeled translation products from hybrid-selected RNA (lane HS) and a control with no RNA (lane NR) are shown. Pulse-labeled proteins from mock-infected cells (lane M) and infected cells at 12 h p.i. (lane 12) are shown. Anti-CG-4- $\beta$ -galactosidase antiserum was used to immunoprecipitate proteins from pulse-labeled lysates of mock-infected and infected cells at 12 h p.i. and from [ $^{35}$ S]methionine-labeled in vitro translation products (lanes c to e). Molecular weights are shown at the left (K,  $10^3$ ). The 42-kDa capsid protein is indicated by an arrow, and the endogenous globin translation product (endog) is noted.

for regulation of the AcMNPV infection (10). Possible AcMNPV proteins that may be required for late transcription include positive or negative regulatory factors and a virus-specific RNA polymerase (12).

There are at least three initiation sites for capsid gene transcription as indicated by both nuclease protection and primer extension analyses. Like other late and very late baculovirus transcripts (17, 19, 47), the two major capsid gene transcripts initiate at the central A of a highly conserved ATAAG sequence located within an A+T-rich region of the genome. It is interesting, however, that the most distal

Lanes A, C, G, and T contain the sequencing reactions. Sizes of molecular markers (lane MW) and the 5'-end-labeled probe (lane P) are shown on the right (the position of the probe is marked with an asterisk). The transcriptional start sites were determined by comparing the protected fragments and the extension products with the sequencing ladder. The sequence around each transcriptional start site is printed to the left of the sequencing ladder; arrows denote the 5' initiation site. The positions of the 5' ends are indicated by solid arrows in the diagram in panel A. The sizes of the major protected fragments and extension products are shown on the far left adjacent to the sequence. (D) Long exposure of a gel analyzing primer extension products and a sequencing ladder from the 15-base primer as described for panel C. The sequencing reactions are in lanes A, C, G, and T. Primer extensions from total RNA isolated from mock-infected cells and AcMNPV-infected cells at 6 and 12 h p.i. are in lanes M, 6, and 12, respectively. The sizes (in nucleotides) of molecular markers (lanes MW) are shown to the right of the panel. The sequences around the start sites are indicated to the left of the panel as described for panel C.



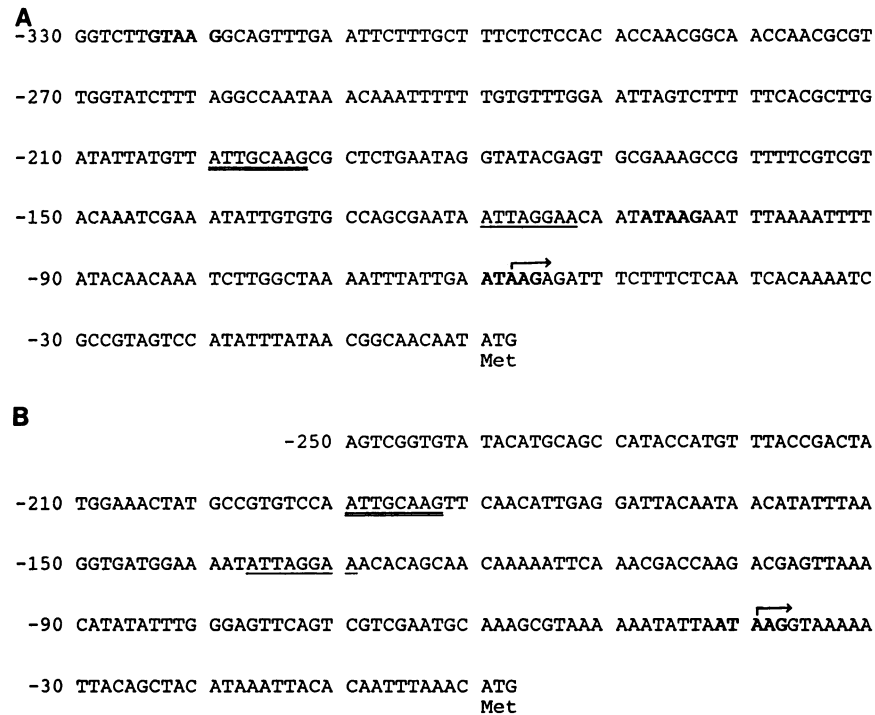


FIG. 7. (A) Sequences upstream of the initial methionine codon of the capsid protein gene. (B) Sequences upstream of the initial methionine codon of the 6.9-kDa basic core protein gene (47). The first nucleotide in each row, counted from the initial ATG, is numbered to the left. The  $\hat{G}$ TAAAG initiation sequences are printed in bold face, and the proximal transcriptional start sites are indicated by arrows. The conserved octamers ATTAGGAA and ATTGCAAG are indicated by single and double underlines, respectively.

and least abundant of the three capsid gene transcripts initiates within a GTAAG sequence. Analysis of the polyhedrin promoter region by linker scan mutational analysis demonstrated that the ATAAG sequence at the polyhedrin transcriptional start site which overlaps with an octanucleotide site, TAAGTATT, is the major determinant of high-level polyhedrin gene expression (32). These studies also revealed that a mutation in which the ATAAG was changed to GTAAG still allowed strong expression, although it was reduced approximately fourfold. The presence of more than one of these  $\hat{G}$ TAAAG motifs in the capsid promoter is likely to be significant. Multiple initiation sites may allow the capsid gene to compete more effectively for RNA polymerase, allowing the message to be transcribed at high levels during BV synthesis (10 to 18 h p.i.) preceding occlusion.

The capsid gene is transcribed into at least three RNA species based on the existence of three different 5' initiation sites. A major transcriptional stop signal for the capsid transcript that terminates closest to the end of the capsid ORF, at approximately nucleotide 1560, is poorly recognized, and the majority of transcripts proceed through a downstream early ORF (Thiem and Miller, submitted), thus generating a bicistronic transcript. As in other regions of the AcMNPV genome that have been examined, there is little space between potential coding regions. In this region, the ATG of the downstream ORF is only two nucleotides from the translational termination signal (TAA) of the capsid gene. Readthrough may be important in the regulation of the downstream ORF.

Both late genes, such as the capsid protein and the basic 6.9-kDa core protein (47) genes, and occlusion-specific genes, such as the polyhedrin (17) and p10 (19) genes, contain the conserved 5' initiation sequence  $\hat{G}$ TAAAG. To

determine whether there were any other features in the promoter regions of these genes that might account for differences in the temporal regulation of late and very late gene expression, we compared the sequences upstream from the ATG for any similarities. The upstream sequences of both the 6.9-kDa protein and the capsid protein genes contained two octamers, ATTAGGAA and ATTGCAAG, that were not observed in the occlusion-specific genes. The sequence ATTAGGAA was located at nucleotide positions -120 in the capsid sequence (Fig. 7A) and -143 in the core sequence (Fig. 7B) with respect to the initiation ATG. The sequence ATTGCAAG was located at nucleotide position -200 (Fig. 7A) in the capsid and -209 (Fig. 7B) in the core. The placement of these sequences with respect to the ATAAG transcriptional start sites may be of significance. The octamer ATTGCAAG (double underline in Fig. 7) is approximately the same distance, 140 bp in the capsid and 148 bp in the 6.9-kDa core protein gene, from the ATAAG proximal to the initial methionine in each gene. Although the spacing of the octamers and the ATAAG transcriptional initiation sites are not identical, the linear arrangement of the octamers and the ATAAG sequences, ATTGCAAG-ATTAGGAA-ATAAG, is conserved between the two genes. The presence of these two octamer sequences leads to the intriguing hypothesis that protein factors could interact specifically at these sites and act either to initiate transcription at earlier times than polyhedrin or p10 or to repress transcription at very late times p.i. This hypothesis can be tested by linker scan mutational analysis.

The region of homology between the rotavirus capsid protein (4) and the AcMNPV capsid protein spanned 80 amino acids, 23% of the AcMNPV capsid protein. The



- nuclear polyhedrosis virus. *J. Invertebr. Pathol.* **36**:431–432.
32. Rankin, C., B. G. Ooi, and L. K. Miller. 1988. Eight base pairs at the transcriptional start point are the major determinant for baculovirus polyhedrin expression. *Gene* **70**:39–49.
  33. Rice, W. C., and L. K. Miller. 1986/87. Baculovirus transcription in the presence of inhibitors and in nonpermissive *Drosophila* cells. *Virus Res.* **6**:155–172.
  34. Roberts, P. L., and W. Naser. 1982. Characterization of monoclonal antibodies to the *Autographa californica* nuclear polyhedrosis virus. *Virology* **122**:424–430.
  35. Sanger, F., S. Nicklen, and A. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463–5467.
  36. Smith, G. E., and M. D. Summers. 1981. Application of a novel radioimmunoassay to identify baculovirus structural proteins that share interspecies antigenic determinants. *J. Virol.* **39**:125–137.
  37. Smith, G. E., J. M. Vlak, and M. D. Summers. 1982. *In vitro* translation of *Autographa californica* nuclear polyhedrosis virus early and late mRNAs. *J. Virol.* **44**:199–208.
  38. Summers, M. D., and G. E. Smith. 1978. Baculovirus structural polypeptides. *Virology* **84**:390–402.
  39. Towbin, H., R. Staehelin, and J. Gordon. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci. USA* **76**:4350–4354.
  40. Tweeten, K. A., L. A. Bulla, Jr., and R. A. Consigli. 1980. Characterization of an extremely basic protein derived from granulosis virus nucleocapsids. *J. Virol.* **33**:866–876.
  41. Vaitukaitis, J. L. 1981. Production of antisera with small doses of immunogen: multiple intradermal injections. *Methods Enzymol.* **73**:46–52.
  42. Vaughn, J. L., R. H. Goodwin, G. L. Tompkins, and P. McCawley. 1977. Establishment of two insect cell lines from the insect *Spodoptera frugiperda* (Lepodoptera: Noctuidae). *In Vitro* **13**:213–217.
  43. Watson, R. J., J. H. Weis, J. S. Salstrom, and L. W. Enquist. 1982. Herpes simplex virus type-1 glycoprotein D gene: nucleotide sequence and expression in *Escherichia coli*. *Science* **218**:381–384.
  44. Weaver, R. F., and C. Weissman. 1979. Mapping of RNA by a modification of the Berk-Sharp procedure: the 5' termini of 15S  $\beta$ -globin mRNA precursor and mature 10S  $\beta$ -globin mRNA have identical map coordinates. *Nucleic Acids Res.* **7**:1175–1193.
  45. Weis, J. H., L. W. Enquist, and R. J. Watson. 1983. An immunologically active chimaeric protein containing herpes simplex type 1 glycoprotein D. *Nature (London)* **302**:72–74.
  46. Wickens, M., and P. Stephenson. 1984. Role of the conserved AAUAAA sequence: four AAUAAA point mutants prevent messenger RNA 3' end formation. *Science* **226**:1045–1051.
  47. Wilson, M. E., T. H. Mainprize, P. D. Friesen, and L. K. Miller. 1987. Location, transcription and sequence of a baculovirus gene encoding a small arginine-rich polypeptide. *J. Virol.* **61**:661–666.